

Update on the Demographic Profile and the Demographic and Housing Characteristics File

CSAC Discussion

Barbara Entwisle, PhD

Kenan Distinguished Professor of Sociology

Fellow, Carolina Population Center

University of North Carolina at Chapel Hill

March 17, 2022

Progress on release of next DHC demonstration products.

DHC “person” tables released on March 16, yesterday.

DHC “housing” tables to be released soon.

114 Person File tables, 119 Housing Unit File tables, 66 geographic levels.

Given the importance of the DHC tables, the amount of detail to be released, and the fact that the informational webinar will occur almost a week after the release of the “person” tables, CSAC recommends a comment period of 45 days, ideally for both sets of tables.

First Look: Net Mean Error for Counties

Three tables reporting net mean error at the county level for broad sex x age groups, 5-year age groups, and household type.

Substantial reduction in net mean error compared to earlier demonstration products, especially for “persons.”

Are there prospects for further improvement for “units?”

CSAC looks forward to the opportunity to inspect measures of accuracy, bias, and outliers associated with the new demonstration products and reserves the right to ask questions at that time.

Comments and Use Cases

Newly released demonstration products will be responsive to comments and use cases submitted last fall, at least in part.

After a careful assessment of cost for privacy and accuracy, the Bureau may not release the detail needed for all desired uses of the data.

CSAC recommends that the Bureau share information about use cases, both those supported by special tabulations and those which are not.

Race, Ethnicity, and Age

Public quite concerned about consequences of undercount in the Decennial for underserved populations.

Likely equal interest in the accuracy, bias, and outliers associated with race, ethnicity, and age (especially young children) in the DP and DH.

Importance of communicating effectively to a variety of different audiences given potential confusion about sources of error in census counts.

Public Use Microdata Sample (PUMS)

CSAC enthusiastically welcomes the news that planning is underway to develop a PUMS based on the 2020 Decennial.

PUMS data are a vital national resource, used extensively in demographic, economic, and social research, and important to policymakers and policymaking.

CSAC recommends that the Bureau share its plans and timetable for a PUMS file as soon as possible.

Reconstruction and Reidentification of DHC

The universe for the reidentification effort is all records in the census edited file (the CEF) with a Protected Identification Key (PIK). While the reasons for defining the universe as all records in the census edited file (CEF) are understandable, what are the consequences of restricting the analysis in this way?

Reconstruction and Reidentification: Slide 29

Different methods and levels of disclosure protection are much less differentiated in their matches to commercial data than to the CEF.

Is it correct to understand the CEF as a proxy for better (more accurate) commercial data?

Is there concern about how closely the reconstruction matches the CEF, apart from its role as a proxy?

Reconstruction and Reidentification: Uniques

Cutting ρ in half buys very little reduction in risk. Does this provide an argument for the larger ρ ?

Thank you